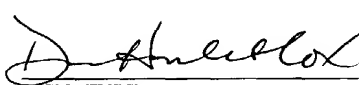


16 JAN 2001

FORM PCT-1390 U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE <b>TRANSMITTAL LETTER TO THE UNITED STATES          DESIGNATED/ELECTED OFFICE (DO/EO/US)          CONCERNING A FILING UNDER 35 U.S.C. 371</b>		ATTORNEY'S DOCKET NUMBER <b>PF-0564 USN</b>
INTERNATIONAL APPLICATION NO. <b>PCT/US99/12385</b>		U.S. APPLICATION NO. (If known, see 37 CFR 1.5) <b>TO BE ASSIGNED</b> <b>09/744197</b> PRIORITY DATE CLAIMED <b>20 July 1998</b>
TITLE OF INVENTION <b>HUMAN CALCIUM REGULATORY PROTEINS</b>		
APPLICANT(S) FOR DO/EO/US <b>INCYTE PHARMACEUTICALS, INC.; BANDMAN, Olga; TANG, Y. Tom; CORLEY, Neil C.; GUEGLER, Karl J.;          BAUGHN, Mariah R.; JUNMING, Yang</b>		
Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information: 1. <input checked="" type="checkbox"/> This is the <b>FIRST</b> submission of items concerning a filing under 35 U.S.C. 371. 2. <input type="checkbox"/> This is a <b>SECOND</b> or <b>SUBSEQUENT</b> submission of items concerning a filing under 35 U.S.C. 371. 3. <input type="checkbox"/> This is an express request to promptly begin national examination procedures (35 U.S.C. 371 (f)). 4. <input type="checkbox"/> The US has been elected by the expiration of 19 months from the priority date (PCT Article 31). 5. <input checked="" type="checkbox"/> A copy of the International Application as filed (35 U.S.C. 371(c)(2)) a. <input type="checkbox"/> is attached hereto (required only if not communicated by the International Bureau) b. <input type="checkbox"/> has been communicated by the International Bureau. c. <input checked="" type="checkbox"/> is not required, as the application was filed in the United States Receiving Office (RO/US). 6. <input type="checkbox"/> An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)). 7. <input checked="" type="checkbox"/> Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3)) a. <input type="checkbox"/> are attached hereto (required only if not communicated by the International Bureau). b. <input type="checkbox"/> have been communicated by the International Bureau. c. <input type="checkbox"/> have not been made; however, the time limit for making such amendments has NOT expired. d. <input checked="" type="checkbox"/> have not been made and will not be made. 8. <input type="checkbox"/> An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)). 9. <input type="checkbox"/> An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)). 10. <input type="checkbox"/> An English language translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).		
<b>Items 11 to 16 below concern document(s) or information included:</b> 11. <input type="checkbox"/> An Information Disclosure Statement under 37 CFR 1.97 and 1.98. 12. <input type="checkbox"/> An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.27 and 3.31 is included. 13. <input type="checkbox"/> A <b>FIRST</b> preliminary amendment. <input type="checkbox"/> A <b>SECOND</b> or <b>SUBSEQUENT</b> preliminary amendment. 14. <input type="checkbox"/> A substitute specification. 15. <input type="checkbox"/> A change of power of attorney and/or address letter. 16. <input checked="" type="checkbox"/> Other items or information:  1) Transmittal Letter (2 pp, in duplicate) 2) Return Postcard 3) Express Mail Label No.: <b>EL 743 378 859 US</b>		

JC07 Rec'd PCT/PTO 16 JAN 2001

U.S. APPLICATION NO. (if known) <b>09/744197</b> TO BE ASSIGNED		INTERNATIONAL APPLICATION NO.: PCT/US99/12385		ATTORNEY'S DOCKET NUMBER PF-0564 USN	
17. <input type="checkbox"/> The following fees are submitted: <b>BASIC NATIONAL FEE (37 CFR 1.492(a)(1)-(5)):</b> Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO..... <b>\$1000.00</b> International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO.. <b>\$860.00</b> International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO..... <b>\$710.00</b> <input checked="" type="checkbox"/> <b>International preliminary examination fee paid to USPTO (37 CFR 1.482)</b> <b>but all claims did not satisfy provisions of PCT Article 33(1)-(4).....\$690.00</b> <input type="checkbox"/> International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(1)-(4).....\$100.00					
ENTER APPROPRIATE BASIC FEE AMOUNT =				\$690.00	
Surcharge of <b>\$130.00</b> for furnishing the oath or declaration later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(c)).				\$	
CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE		
Total Claims	21 =	1	X \$ 18.00	\$ 18.00	
Independent Claims	2 =	0	X \$ 80.00	\$	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ \$270.00	\$	
TOTAL OF ABOVE CALCULATIONS =				\$708.00	
<input type="checkbox"/> Applicant claims small entity status. Sec 37 CFR 1.27. The fees indicated above are reduced by 1/2.				\$	
SUBTOTAL =				\$708.00	
Processing fee of <b>\$130.00</b> for furnishing the English translation later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(f)).				\$	
TOTAL NATIONAL FEE =				\$708.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by the appropriate cover sheet (37 CFR 3.28, 3.31). <b>\$40.00</b> per property				\$	
TOTAL FEES ENCLOSED =				\$708.00	
				Amount to be Refunded:	\$
				Charged:	\$
a. <input type="checkbox"/> A check in the amount of \$ _____ to cover the above fees is enclosed. b. <input checked="" type="checkbox"/> Please charge my Deposit Account No. <u>09-0108</u> in the amount of \$ <u>708.00</u> to cover the above fees. c. <input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. <u>09-0108</u> . A duplicate copy of this sheet is enclosed.  <b>NOTE:</b> Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.  SEND ALL CORRESPONDENCE TO:  INCYTE GENOMICS, INC. 3160 Porter Drive Palo Alto, CA 94304  <div style="text-align: center;">           SIGNATURE           NAME: Diana Hamlet-Cox           REGISTRATION NUMBER: 33,302           DATE: <u>16</u> January 2001       </div>					

HUMAN CALCIUM REGULATORY PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of calcium regulatory proteins and to the use of these sequences in the diagnosis, treatment, and prevention of neurological and cardiovascular disorders and cancer.

BACKGROUND OF THE INVENTION

In nearly all eukaryotic cells, calcium ( $\text{Ca}^{2+}$ ) functions as an intracellular signaling molecule in diverse cellular processes including cell proliferation, neurotransmitter secretion, glycogen metabolism, and muscle contraction. Within a resting cell, the cytosolic concentration of  $\text{Ca}^{2+}$  is less than  $10^{-7}$  M. However, when the cell is stimulated by an external signal, such as a neural impulse or a growth factor, the cytosolic concentration of  $\text{Ca}^{2+}$  increases by about 50-fold. This influx of  $\text{Ca}^{2+}$  is prompted by the opening of plasma membrane  $\text{Ca}^{2+}$  channels and by the release of  $\text{Ca}^{2+}$  from intracellular stores such as the endoplasmic reticulum.  $\text{Ca}^{2+}$  directly activates regulatory enzymes, such as protein kinase C, which trigger signal transduction pathways.  $\text{Ca}^{2+}$  also binds to specific  $\text{Ca}^{2+}$  binding proteins which then activate multiple target proteins including enzymes, membrane transport pumps, and ion channels.

Some  $\text{Ca}^{2+}$  binding proteins are characterized by the presence of one or more EF-hand  $\text{Ca}^{2+}$  binding motifs. (Reviewed in Celio, M. R. et al. (1996) Guidebook to the Calcium-Binding Proteins, Oxford University Press, New York, NY, pp. 15-20.) The EF-hand motif is comprised of 12 amino acids flanked by  $\alpha$ -helices. Acidic residues are generally conserved at positions 1, 3, 5, 7, 9, and 12, and glycine is often conserved at position 6. At least 250 EF-hand proteins defining 39 families have been identified in various tissues. For example, S100 $\alpha$  is a member of the S100 protein family isolated from human heart. (Celio et al. supra, pp. 135-136; Engelkamp, D. et al. (1992) Biochemistry 31:10258-10264.) S100 $\alpha$  messenger RNA expression is restricted to the heart, skeletal muscle, and brain. Normally, S100 $\alpha$  is undetectable in serum; however, S100 $\alpha$  is detectable in serum of patients with renal carcinoma at levels correlated with clinical progression of the disease. In addition, S100 $\alpha$  levels are increased in serum following myocardial infarction.

Other  $\text{Ca}^{2+}$  binding proteins include those involved in  $\text{Ca}^{2+}$  sequestration and protein folding in the endoplasmic reticulum (ER). For example,  $\text{Ca}^{2+}$  binding proteins 1 and 2 (CaBP1 and CaBP2) are protein disulfide isomerases that contain two and three thioredoxin-like active

sites, respectively. (Fullekrug, J. et al. (1994) J. Cell Sci. (1994) 107:2719-2727; Van, P. N. et al. (1993) Eur. J. Biochem. 213:789-795.) Each active site contains two invariant cysteines which directly participate in reversible oxidation/reduction reactions. CaBP1 and CaBP2 may facilitate the folding of nascent proteins in the ER by catalyzing the formation and isomerization of disulfide bonds. Like all resident soluble ER proteins, CaBP1 and CaBP2 each contain the C-terminal KDEL/KEEL tetrapeptide required for their retention in the ER.

Other  $\text{Ca}^{2+}$  binding proteins include the annexins which comprise a ubiquitous family of  $\text{Ca}^{2+}$ -dependent phospholipid binding proteins. (Celio et al. supra, pp. 181-184.) There are eleven structurally distinct classes of annexins characterized by a variable number of core domain repeats. Although the precise function of the annexins are unknown, evidence suggests that annexins are involved in intracellular signaling, vesicular transport, and cytoskeletal function.

Regulation of intracellular  $\text{Ca}^{2+}$  levels by  $\text{Ca}^{2+}$  binding proteins is critical for neural signaling and muscle contraction. In particular, elevated levels of intracellular calcium are a major cause of cardiovascular dysfunction, including myocardial ischemia leading to acute infarction.

Current therapies for the treatment of myocardial ischemia block the influx of calcium into myocardial and vascular smooth muscle cells.

The discovery of new  $\text{Ca}^{2+}$  regulatory proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, treatment, and prevention of neurological and cardiovascular disorders and cancer.

## SUMMARY OF THE INVENTION

The invention is based on the discovery of new  $\text{Ca}^{2+}$  regulatory proteins (CaREG), the polynucleotides encoding CaREG, and the use of these compositions for the diagnosis, treatment, or prevention of neurological and cardiovascular disorders and cancer.

The invention features substantially purified polypeptides,  $\text{Ca}^{2+}$  regulatory proteins, referred to collectively as "CaREG" and individually as "CaREG-1" and "CaREG-2." In one aspect, the invention provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, and fragments thereof.

The invention further provides a substantially purified variant having at least 90% amino acid identity to the amino acid sequence of SEQ ID NO:1, SEQ ID NO:2, or to a fragment of any of these sequences. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, and fragments thereof. The invention also includes an isolated and purified



NO:2, and fragments thereof, as well as a purified agonist and a purified antagonist to the polypeptide.

The invention also provides a method for treating or preventing a neurological disorder, the method comprising administering to a subject in need of such treatment an effective amount of  
5 a pharmaceutical composition comprising a substantially purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, and fragments thereof.

The invention also provides a method for treating or preventing a cardiovascular disorder, the method comprising administering to a subject in need of such treatment an effective amount of  
10 a pharmaceutical composition comprising a substantially purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, and fragments thereof.

The invention also provides a method for treating or preventing a cancer, the method comprising administering to a subject in need of such treatment an effective amount of an  
15 antagonist of the polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, and fragments thereof.

The invention also provides a method for detecting a polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, and fragments thereof in a biological sample containing nucleic acids, the  
20 method comprising the steps of: (a) hybridizing the complement of the polynucleotide sequence encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, and fragments thereof to at least one of the nucleic acids of the biological sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a  
25 polynucleotide encoding the polypeptide in the biological sample. In one aspect, the method further comprises amplifying the polynucleotide prior to hybridization.

## BRIEF DESCRIPTION OF THE FIGURES

Figures 1A and 1B show the amino acid sequence (SEQ ID NO:1) and nucleic acid  
30 sequence (SEQ ID NO:3) of CaREG-1. The alignment was produced using MACDNASIS PRO software (Hitachi Software Engineering Co. Ltd., Yokohama, Japan).

Figures 2A, 2B, 2C, 2D, 2E, 2F, and 2G show the amino acid sequence (SEQ ID NO:2) and nucleic acid sequence (SEQ ID NO:4) of CaREG-2. The alignment was produced using MACDNASIS PRO software.

Table 1 shows the programs, their descriptions, references, and threshold parameters (where appropriate) used to analyze CaREG.

## DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## DEFINITIONS

"CaREG," as used herein, refers to the amino acid sequences, or variant thereof, of substantially purified CaREG obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist," as used herein, refers to a molecule which, when bound to CaREG, increases or prolongs the duration of the effect of CaREG. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of CaREG.

An "allelic variant," as this term is used herein, is an alternative form of the gene encoding CaREG. Allelic variants may result from at least one mutation in the nucleic acid sequence and

may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms.

Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur

5 alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding CaREG, as described herein, include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same as CaREG or a polypeptide with at least one functional characteristic of CaREG. Included within this definition are polymorphisms which may or may not be readily

10 detectable using a particular oligonucleotide probe of the polynucleotide encoding CaREG, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding CaREG. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent CaREG. Deliberate amino  
15 acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of CaREG is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar  
20 hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms "amino acid" or "amino acid sequence," as used herein, refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments," "immunogenic  
25 fragments," or "antigenic fragments" refer to fragments of CaREG which are preferably at least 5 to about 15 amino acids in length, most preferably at least 14 amino acids, and which retain some biological activity or immunological activity of CaREG. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native  
30 amino acid sequence associated with the recited protein molecule.

"Amplification," as used herein, relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art. (See, e.g., Dieffenbach, C.W. and G.S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, pp.1-5.)



The term "antagonist," as it is used herein, refers to a molecule which, when bound to CaREG, decreases the amount or the duration of the effect of the biological or immunological activity of CaREG. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of CaREG.

5 As used herein, the term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind CaREG polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be  
10 derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant," as used herein, refers to that fragment of a molecule  
15 (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

20 The term "antisense," as used herein, refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation  
25 "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

As used herein, the term "biologically active," refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic CaREG,  
30 or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity," as used herein, refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" binds to the complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules





contain all of the elements required for stable mitotic chromosome segregation and maintenance.  
(See, e.g., Harrington, J.J. et al. (1997) Nat Genet. 15:345-355.)

The term "humanized antibody," as used herein, refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more  
5 closely resembles a human antibody, and still retains its original binding ability.

"Hybridization," as the term is used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

As used herein, the term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary  
10 bases. A hybridization complex may be formed in solution (e.g.,  $C_0t$  or  $R_0t$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" or "addition," as used herein, refer to changes in an amino acid or  
15 nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which  
20 may affect cellular and systemic defense systems.

The term "microarray," as used herein, refers to an arrangement of distinct polynucleotides arrayed on a substrate, e.g., paper, nylon or any other type of membrane, filter, chip, glass slide, or any other suitable solid support.

The terms "element" or "array element" as used herein in a microarray context, refer to  
25 hybridizable polynucleotides arranged on the surface of a substrate.

The term "modulate," as it appears herein, refers to a change in the activity of CaREG. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of CaREG.

The phrases "nucleic acid" or "nucleic acid sequence," as used herein, refer to a  
30 nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material. In this context, "fragments" refers to those nucleic acid sequences which, comprise a region of unique polynucleotide sequence that specifically identifies







D111 and from E115 to L128 are similar to an annexin type I signature. CaREG-1 has chemical and structural similarity with human S100 $\alpha$  (GI 36176). In particular, the region of CaREG-1 from S54 to K147 shares 26% amino acid identity with S100 $\alpha$ . A fragment of SEQ ID NO:3 from about nucleotide 25 to about nucleotide 54 is useful as a hybridization probe. Northern analysis shows that this sequence is expressed exclusively in cDNA libraries derived from lung tissue, including fetal and cancerous lung tissue.

Nucleic acids encoding the CaREG-2 of the present invention were first identified in Incyte Clone 2176876 from the dermal microvascular endothelial cell cDNA library (ENDCNOT03) using a computer search, e.g., BLAST, for amino acid sequence alignments. A consensus sequence, SEQ ID NO:4, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 2176876H1 (ENDCNOT03), 547705X58C1 (BEPINOT01), 1553296F1 (BLADTUT04), 1417868F1 (KIDNNOT09), 1814186F6 (PROSNOT20), 1310941F1 (COLNFET02), 1225462H1 (COLNTUT02), 001150R1 (U937NOT01), and shotgun sequence SADA00307R1.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:2, as shown in Figures 2A, 2B, 2C, 2D, 2E, 2F, and 2G. CaREG-2 is 363 amino acids in length and has one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at S340; ten potential casein kinase II phosphorylation sites at T29, S39, T98, S114, S128, S216, T237, T266, T290, and S295; five potential protein kinase C phosphorylation sites at T69, T170, S216, S295, and S323; and one potential tyrosine kinase phosphorylation site at Y82. BLOCKS, PRINTS, PFAM, and PROFILESCAN all indicate that the complete amino acid sequence of CaREG-2 shows extensive similarity to thioredoxin protein signatures. In particular, three thioredoxin consensus active sites are conserved in CaREG-2 from F12 to W30, from F140 to W158, and from F273 to W291. Included within each active site region are invariant cysteine pairs at C20 and C23, C148 and C151, and C281 and C284. A predicted signal peptide from M1 to C23 and an ER retention signal from K360 to L363 suggests that CaREG-2 is a resident ER protein. CaREG-2 has chemical and structural similarity with regions of rat CaBP2 (GI 393203). In particular, CaREG-2 shares 26% overall amino acid identity with CaBP2. A fragment of SEQ ID NO:4 from about nucleotide 163 to about nucleotide 192 is useful as a hybridization probe. Northern analysis shows the expression of this sequence in various cDNA libraries, at least 63% of which are associated with cancer or proliferating cells and at least 37% of which are associated with the immune response or trauma. Northern analysis further shows that 27% of the libraries expressing CaREG-2 are derived from reproductive tissue, and 17% from cardiovascular tissue. In addition, the highest levels of CaREG-2 expression are observed in libraries derived from





CaREG derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding CaREG or any fragment thereof.

5 Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:3, SEQ ID NO:4, and fragments thereof, under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) For example, stringent salt concentration will ordinarily be less  
10 than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature  
15 conditions will ordinarily include temperatures of at least about 30°C, more preferably of at least about 37°C, and most preferably of at least about 42°C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred  
20 embodiment, hybridization will occur at 30°C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37°C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42°C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50 % formamide, and 200 µg/ml ssDNA. Useful variations on  
25 these conditions will be readily apparent to those skilled in the art.

The washing steps which follow hybridization can also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For  
example, stringent salt concentration for the wash steps will preferably be less than about 30 mM  
30 NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperature of at least about 25°C, more preferably of at least about 42°C, and most preferably of at least about 68°C. In a preferred embodiment, wash steps will occur at 25°C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at

42°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS.

Additional variations on these conditions will be readily apparent to those skilled in the art.

Methods for DNA sequencing and analysis are well known in the art. The methods may  
 5 employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE enzyme (Amersham International PLC, Buckinghamshire England), TAQ polymerase (Perkin-Elmer Corp.), thermostable T7 polymerase (Amersham International PLC), or combinations of polymerases and proofreading exonucleases, such as those found in the ELONGASE amplification system (Life Technologies, Inc., Gaithersburg, MD). Preferably, sequence preparation is  
 10 automated with machines, e.g., the ABI CATALYST 800 (Perkin-Elmer Corp.) or MICROLAB 2200 systems (Hamilton Co., Reno, NV), in combination with thermal cyclers. Sequencing can also be automated, such as by ABI PRISM 373 or 377 systems (Perkin-Elmer Corp.) or the MEGABACE 1000 capillary electrophoresis system (Molecular Dynamics, Inc., Sunnyvale, CA). Sequences can be analyzed using computer programs and algorithms well known in the art. (See,  
 15 e.g., Ausubel, et al. (1997) Short Protocols in Molecular Biology, Wiley, New York, NY, unit 7.7; and Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, Inc, New York, NY.)

The nucleic acid sequences encoding CaREG may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect  
 20 upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived  
 25 from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an  
 30 engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-306). Additionally, one may use PCR, nested primers, and PROMOTORFINDER libraries (Clontech, Palo Alto, CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon

junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO™ 4.06 Primer Analysis software (National Biosciences Inc., Plymouth, MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to  
5 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of  
10 sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for  
15 detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR software, (The Perkin-Elmer Corp., Norwalk, CT)), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited  
20 amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode CaREG may be cloned in recombinant DNA molecules that direct expression of CaREG, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the  
25 same or a functionally equivalent amino acid sequence may be produced and used to express CaREG.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter CaREG-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the  
30 gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In another embodiment, sequences encoding CaREG may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, and Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232.) Alternatively, CaREG itself or a fragment thereof may be synthesized using chemical methods.

5 For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A Peptide Synthesizer (The Perkin-Elmer Corp., Norwalk, CT). Additionally, the amino acid sequence of CaREG, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant  
10 polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH  
15 Freeman and Co., New York, NY.)

In order to express a biologically active CaREG, the nucleotide sequences encoding CaREG or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as  
20 enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding CaREG. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding CaREG. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding CaREG and its initiation  
25 codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic.  
30 The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding CaREG and appropriate transcriptional and

translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, ch. 4, 8, and 16-17; and Ausubel, F.M. et al. (1995, and periodic supplements) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding CaREG. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus (CaMV) or tobacco mosaic virus (TMV)) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding CaREG. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding CaREG can be achieved using a multifunctional E. coli vector such as pBLUESCRIPT plasmid (Stratagene) or pSPORT1 plasmid (Life Technologies, Inc.). Ligation of sequences encoding CaREG into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of CaREG are needed, e.g. for the production of antibodies, vectors which direct high level expression of CaREG may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of CaREG. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, supra; and Grant et al. (1987) Methods Enzymol. 153:516-54; Scorer, C. A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of CaREG. Transcription of sequences

encoding CaREG may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV. (Takamatsu, N. (1987) EMBO J. 6:307-311.) Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., Hobbs, S. or Murry, L.E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, NY; pp. 191-196.)

10 In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding CaREG may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses CaREG in host cells. (See, e.g., Logan, J. and T. Shenk 15 (1984) Proc. Natl. Acad. Sci. 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb 20 are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

For long term production of recombinant proteins in mammalian systems, stable expression of CaREG in cell lines is preferred. For example, sequences encoding CaREG can be transformed into cell lines using expression vectors which may contain viral origins of replication 25 and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed 30 cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* or *apv* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; and Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite,

antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* or *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-3570;

- 5 Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14; and Murry, supra.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP) (Clontech, Palo Alto, CA),  $\beta$  glucuronidase and its substrate  $\beta$ -D-glucuronoside, or luciferase and its substrate luciferin may be
- 10 used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. et al. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For

15 example, if the sequence encoding CaREG is inserted within a marker gene sequence, transformed cells containing sequences encoding CaREG can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding CaREG under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

- 20 In general, host cells that contain the nucleic acid sequence encoding CaREG and that express CaREG may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or
- 25 protein sequences.

Immunological methods for detecting and measuring the expression of CaREG using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay

30 utilizing monoclonal antibodies reactive to two non-interfering epitopes on CaREG is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN, Section IV; Coligan, J. E. et al. (1997 and periodic supplements) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York, NY; and Maddox,



D.E. et al. (1983) J. Exp. Med. 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding CaREG  
5 include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding CaREG, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures  
10 may be conducted using a variety of commercially available kits, such as those provided by Pharmacia & Upjohn (Kalamazoo, MI), Promega (Madison, WI), and U.S. Biochemical Corp. (Cleveland, OH). Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

15 Host cells transformed with nucleotide sequences encoding CaREG may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CaREG may be designed to contain signal  
20 sequences which direct secretion of CaREG through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro"  
25 form of the protein may also be used to specify protein targeting, folding, and/or activity.

Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Bethesda, MD) and may be chosen to ensure the correct modification and processing of the foreign protein.

30 In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding CaREG may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric CaREG protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of CaREG activity.

Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His  
 5 enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the  
 10 CaREG encoding sequence and the heterologous protein sequence, so that CaREG may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel, F. M. et al. (1995 and periodic supplements) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, ch 10. A variety of commercially available kits may also be used to facilitate expression and purification of fusion  
 15 proteins.

In a further embodiment of the invention, synthesis of radiolabeled CaREG may be achieved in vitro using the TNT™ rabbit reticulocyte lysate or wheat germ extract systems (Promega, Madison, WI). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the  
 20 presence of a radiolabeled amino acid precursor, preferably <sup>35</sup>S-methionine.

Fragments of CaREG may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the Applied Biosystems 431A peptide synthesizer (Perkin-  
 25 Elmer Corp.). Various fragments of CaREG may be synthesized separately and then combined to produce the full length molecule.

## THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists  
 30 between regions of CaREG-1 and Ca<sup>2+</sup> binding proteins such as human S100α (GI 36176) and annexin type I. CaREG-1 is expressed in fetal and cancerous lung tissue. In addition, chemical and structural similarity exists among regions of CaREG-2, rat CaBP2 (GI 393203), and thioredoxin. CaREG-2 is expressed in cancerous and proliferating tissue and in cardiovascular tissue. Therefore, CaREG appears to play a role in neurological and cardiovascular disorders and

cancer.

Therefore, in one embodiment, CaREG or a fragment or derivative thereof may be administered to a subject to treat or prevent a neurological disorder. Such disorders can include, but are not limited to, epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including anxiety and schizophrenia, akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder.

In another embodiment, a vector capable of expressing CaREG or a fragment or derivative thereof may be administered to a subject to treat or prevent a neurological disorder including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified CaREG in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a neurological disorder including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of CaREG may be administered to a subject to treat or prevent a neurological disorder including, but not limited to, those listed above.

In another embodiment, CaREG or a fragment or derivative thereof may be administered to a subject to treat or prevent a cardiovascular disorder. Such disorders can include, but are not limited to, cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, myocardial ischemia and infarction, myocardial failure, tachyarrhythmia, long QT syndrome, and myotonic muscular

dystrophy.

In an additional embodiment, a vector capable of expressing CaREG or a fragment or derivative thereof may be administered to a subject to treat or prevent a cardiovascular disorder including, but not limited to, those described above.

5 In a further embodiment, a pharmaceutical composition comprising a substantially purified CaREG in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a cardiovascular disorder including, but not limited to, those provided above.

10 In still another embodiment, an agonist which modulates the activity of CaREG may be administered to a subject to treat or prevent a cardiovascular disorder including, but not limited to, those listed above.

In a further embodiment, an antagonist of CaREG may be administered to a subject to treat or prevent a cancer. Such cancers may include, but are not limited to, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers  
15 of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In one aspect, an antibody which specifically binds CaREG may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which  
20 express CaREG.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding CaREG may be administered to a subject to treat or prevent a cancer including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists,  
25 complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to  
30 achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of CaREG may be produced using methods which are generally known in the art. In particular, purified CaREG may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind CaREG. Antibodies to CaREG

may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

5 For the production of polyclonal antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with CaREG or with any fragment or oligopeptide thereof which has immunogenic properties. Rats and mice are preferred hosts for downstream applications involving monoclonal antibody production. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants  
10 include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable. (For review of methods for antibody production and analysis, see, e.g., Harlow, E. and Lane, D. (1988) Antibodies: A Laboratory  
15 Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.)

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to CaREG have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 14 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and  
20 contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of CaREG amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to CaREG may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but  
25 are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci.* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the  
30 splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci.* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce

CaREG-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D.R. (1991) Proc. Natl. Acad. Sci. 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; and Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for CaREG may also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity and minimal cross-reactivity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between CaREG and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CaREG epitopes is preferred, but a competitive binding assay may also be employed. (Maddox, supra.)

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for CaREG. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of CaREG-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple CaREG epitopes, represents the average affinity, or avidity, of the antibodies for CaREG. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular CaREG epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the CaREG-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of CaREG, preferably in active form, from the antibody. (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington, D. C.; and Liddell, J. E. and

















diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including anxiety and schizophrenia, akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; a cardiovascular disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, myocardial ischemia and infarction, myocardial failure, tachyarrhythmia, long QT syndrome, and myotonic muscular dystrophy; and a cancer such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding CaREG may be used in Southern or Northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and ELISA assays; and in microarrays utilizing fluids or tissues from patients to detect altered CaREG expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding CaREG may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding CaREG may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding CaREG in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of CaREG, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding CaREG, under conditions suitable for hybridization or

WO 00/05368

PCT/US99/12385

amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from  
5 standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period  
10 ranging from several days to months.

With respect to cancer, the presence of a high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures  
15 or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding CaREG may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a  
20 polynucleotide encoding CaREG, or a fragment of a polynucleotide complementary to the polynucleotide encoding CaREG, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of CaREG include  
25 radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; and Duplaa, C. et al. (1993) Anal. Biochem. 229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or  
30 colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to

determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding CaREG may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, R.A. (ed.) Molecular Biology and Biotechnology, VCH Publishers New York, NY, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding CaREG on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal



WO 00/05368

PCT/US99/12385

location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, CaREG, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between CaREG and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with CaREG, or fragments thereof, and washed. Bound CaREG is then detected by methods well known in the art. Purified CaREG can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding CaREG specifically compete with a test compound for binding CaREG. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with CaREG.

In additional embodiments, the nucleotide sequences which encode CaREG may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. [Attorney Docket No. PF-0564 P], filed July 20, 1998, are hereby expressly incorporated by reference.

## EXAMPLES

### I. cDNA Library Construction

The LUNGN0T09 cDNA library was constructed using RNA isolated from the lung tissue of a 23-week-old Caucasian male fetus. The pregnancy was terminated following a



Biotech, Ltd. were used. Reading frames were determined using standard methods (Ausubel, supra). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA, extension, and shotgun sequencing  
5 were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 1 summarizes the software programs, their descriptions, references, and threshold parameters where applicable. The references cited in the third column of Table 1 are incorporated by reference herein. Sequences were also analyzed using MACDNASIS PRO software (Hitachi Software Engineering Co., Ltd. San Bruno, CA) and  
10 LASERGENE™ software (DNASTAR).

The polynucleotide sequences were validated by removing vector, linker, and polyA tail sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as GenBank primate, rodent, mammalian,  
15 vertebrate, and eukaryote databases, and BLOCKS to acquire annotation, using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. This was followed by translation of the full length polynucleotide sequences to derive the corresponding full  
20 length amino acid sequences. These full length polynucleotide and amino acid sequences were subsequently analyzed by querying against databases such as the GenBank databases described above and SwissProt, BLOCKS, PRINTS, PFAM, and Prosite.

#### IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a  
25 gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; and Ausubel, supra, ch. 4 and 16.)

Electronic northern analysis was produced using analogous computer techniques. These techniques apply BLAST to search for identical or related molecules in nucleotide databases such  
30 as GenBank or LIFESEQ database (Incyte Pharmaceuticals). The sensitivity of the computer search was modified to determine the specificity of the match. The basis of the search is the product score, which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

$$100$$





appropriate clones were selected, ligated into plasmid, and sequenced.

In like manner, the nucleotide sequences of SEQ ID NO:3 and SEQ ID NO:4 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for 5' extension, and an appropriate genomic library.

5

## **VI. Labeling and Use of Individual Hybridization Probes**

Hybridization probes derived from SEQ ID NO:3 and SEQ ID NO:4 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger  
10 nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ - $^{32}$ P] adenosine triphosphate (Amersham, Chicago, IL), and T4 polynucleotide kinase (DuPont NEN<sup>®</sup>, Boston, MA). The labeled oligonucleotides are substantially purified using a Sephadex<sup>™</sup> G-25 superfine size exclusion dextran bead column (Pharmacia & Upjohn,  
15 Kalamazoo, MI). An aliquot containing  $10^7$  counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN, Boston, MA).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon  
20 membranes (Nytran Plus, Schleicher & Schuell, Durham, NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR<sup>™</sup> film (Kodak, Rochester, NY) is exposed to the blots to film for several hours, hybridization patterns are compared visually.

25

## **VII. Microarrays**

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using  
30 thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the

scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE software. Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) *Science* 270:467-470; and Shalon, D. et al. (1996) *Genome Res.* 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

### VIII. Complementary Polynucleotides

Sequences complementary to the CaREG-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring CaREG. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO™ 4.06 software and the coding sequence of CaREG. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the CaREG-encoding transcript.

### IX. Expression of CaREG

Expression and purification of CaREG is achieved using bacterial or virus-based expression systems. For expression of CaREG in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express CaREG upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of CaREG in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding CaREG by either homologous recombination





mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and PCR 3.1 (Invitrogen, Carlsbad, CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP) (Clontech, Palo Alto, CA), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP, and to evaluate properties, for example, their apoptotic state. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow Cytometry, Oxford, New York, NY.

The influence of CaREG on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding CaREG and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success, NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding CaREG and other genes of interest can be analyzed by Northern analysis or microarray techniques.

## **XII. Production of CaREG Specific Antibodies**

CaREG substantially purified using polyacrylamide gel electrophoresis (PAGE)(see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the CaREG amino acid sequence is analyzed using LASERGENE™

software (DNASTAR Inc.) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel supra, ch. 11.)

- 5 Typically, oligopeptides 15 residues in length are synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry and coupled to KLH (Sigma, St. Louis, MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for
- 10 antipeptide activity by, for example, binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

### **XIII. Purification of Naturally Occurring CaREG Using Specific Antibodies**

- Naturally occurring or recombinant CaREG is substantially purified by immunoaffinity
- 15 chromatography using antibodies specific for CaREG. An immunoaffinity column is constructed by covalently coupling anti-CaREG antibody to an activated chromatographic resin, such as CNBr-activated Sepharose (Pharmacia & Upjohn). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

- Media containing CaREG are passed over the immunoaffinity column, and the column is
- 20 washed under conditions that allow the preferential absorbance of CaREG (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/CaREG binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and CaREG is collected.

### **25 XIV. Identification of Molecules Which Interact with CaREG**

- CaREG, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled CaREG, washed, and any wells with labeled CaREG complex are assayed. Data obtained using different concentrations of
- 30 CaREG are used to calculate values for the number, affinity, and association of CaREG with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred

WO 00/05368

PCT/US99/12385

embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.0E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS and PRINTS databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991; J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and Probability value= 1.0E-3 or less
PFAM	A Hidden Markov Models-based application useful for protein family search.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits, depending on individual protein families

Table 1 cont.

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Score= 4.0 or greater
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. <u>supra</u> ; Wisconsin Package Program Manual, version 9, page M51-59. Genetics Computer Group, Madison, WI.	

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, and fragments thereof.
- 5 2. A substantially purified variant having at least 90% amino acid identity to the amino acid sequence of claim 1.
3. An isolated and purified polynucleotide encoding the polypeptide of claim 1.
- 10 4. An isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide of claim 3.
5. An isolated and purified polynucleotide which hybridizes under stringent  
15 conditions to the polynucleotide of claim 3.
6. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide sequence of claim 3.
- 20 7. An isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:3, SEQ ID NO:4, and fragments thereof.
8. An isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide of claim 7.
- 25 9. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 7.
10. An expression vector comprising at least a fragment of the polynucleotide of  
30 claim 3.
11. A host cell comprising the expression vector of claim 10.
12. A method for producing a polypeptide, the method comprising the steps of:

- a) culturing the host cell of claim 11 under conditions suitable for the expression of the polypeptide; and
- b) recovering the polypeptide from the host cell culture.

5           13.     A pharmaceutical composition comprising the polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.

          14.     A purified antibody which specifically binds to the polypeptide of claim 1.

10          15.     A purified agonist of the polypeptide of claim 1.

          16.     A purified antagonist of the polypeptide of claim 1.

          17.     A method for treating or preventing a neurological disorder, the method  
15 comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 13.

          18.     A method for treating or preventing a cardiovascular disorder, the method  
comprising administering to a subject in need of such treatment an effective amount of the  
20 pharmaceutical composition of claim 13.

          19.     A method for treating or preventing a cancer, the method comprising  
administering to a subject in need of such treatment an effective amount of the antagonist of claim  
16.

25          20.     A method for detecting a polynucleotide in a biological sample, the method comprising the steps of:

- (a) hybridizing the polynucleotide of claim 6 to at least one of the nucleic acids in the biological sample, thereby forming a hybridization complex; and
- 30          (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the polynucleotide encoding the polypeptide in the biological sample.

          21.     The method of claim 20 further comprising amplifying the polynucleotide prior to hybridization.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>7</sup> : C12N 15/12, C07K 14/47, C12N 5/10, C07K 16/18, A61K 38/17, C12Q 1/68		A1	(11) International Publication Number: <b>WO 00/05368</b>
			(43) International Publication Date: 3 February 2000 (03.02.00)
(21) International Application Number: PCT/US99/12385		(US). JUNMING, Yang [CN/US]; 7136 Clarendon Street, San Jose, CA 95129 (US).	
(22) International Filing Date: 19 July 1999 (19.07.99)		(74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).	
(30) Priority Data: 60/160,076 20 July 1998 (20.07.98) US		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US 09/119,547 (CIP) Filed on 20 July 1998 (20.07.98)			
(71) Applicant (for all designated States except US): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).			
(72) Inventors; and (75) Inventors/Applicants (for US only): BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577		Published With international search report.	
(54) Title: HUMAN CALCIUM REGULATORY PROTEINS			
(57) Abstract  The invention provides human calcium regulatory proteins (CaREG) and polynucleotides which identify and encode CaREG. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating or preventing disorders associated with expression of CaREG.			



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

5' GAG AGA ATG GGT GGC TGT ATG CAC AGT ACC CAG GAC AAA TCT CTC CAC TTG GAA 54  
M G G C M H S T Q D K S L H L E

63 72 81 90 99 108  
GGA GAT CCC AAT CCT TCT GCA GCC CCA ACA TCC ACC TGC GCA CCT AGG AAA ATG  
G D P N P S A A P T S T C A P R K M

117 126 135 144 153 162  
CCC AAA AGG ATT TCA ATA TCC AAA CAA CTG GCT TCA GTG AAA GCT CTG AGG AAG  
P K R I S I S K Q L A S V K A L R K

171 180 189 198 207 216  
TGC TCA GAT CTG GAA AAA GCT ATT GCC ACC ACT GCT CTG ATT TTC AGA AAT TCT  
C S D L E K A I A T T A L I F R N S

225 234 243 252 261 270  
TCT GAC TCT GAT GGT AAA CTT GAA AAA GCT ATT GCC AAA GAT CTG CTG CAA ACC  
S D S D G K L E K A I A A K D L L Q T

279 288 297 306 315 324  
CAA TTT AGG AAT TTC GCA GAG GGA CAA GAA ACC AAG CCA AAA TAC AGA GAG ATC  
Q F R N F A E G Q E T K P K Y R E I

333 342 351 360 369 378  
CTT TCT GAA CTT GAT GAG CAC ACA GAA AAT AAG CTA GAT TTT GAA GAC TTC ATG  
L S E L D E H T E N K L L D F E D F M

FIGURE 1A

387 396 405 414 423 432  
ATC TTG CTC TTA AGC ATC ACT GTC ATG TCA GAT CTG CTA CAA AAT ATA CGG AAT  
I L L S I T V M S D L L Q N I R N  
441 450 459 468 477 486  
GTA AAA ATT ATG AAA TGA ACA GTT TTA AAT ATG CTG TAT AAA ATA ATG GCA AAA  
V K I M K  
495 504 513 522 531 540  
GAC AGT GTT ATT AAA ATG TTT CCA TCT TAT TTT TGA TTA ATT GAA TAT ATC TAT  
549 558 567 576 585 594  
CAT GCA TCT GAC ATT GCC TAG GAT GAT TCT TCT GAT TGC TGG TAT TCA GAT CCA ATG  
603 612  
TAA CTC CAA ATA TTT ACC C 3'

FIGURE 1B

5' CCC ACG CGT CCG CGG ACG CTG GGC CAT GAC GGC GAG GAC GGA CAG GAC CCG CAC 54  
 9 18 27 36 45  
 63 72 81 90 99 108  
 AGC AAG CAC CTG TAC ACG GCC GAC ATG TTC ACG CAC GGG ATC CAG AGC GCC CCG  
 M F T H G I Q S A A  
 117 126 135 144 153 162  
 CAC TTC GTC ATG TTC TTC GCG CCC TGG TGT GGA CAC TGC CAG CCG CTG CAG CCG  
 H F V M F F A P W C G H C Q R L Q P  
 171 180 189 198 207 216  
 ACT TGG AAT GAC CTG GGA GAC AAA TAC AAC AGC ATG GAA GAT GCC AAA GTC TAT  
 T W N D L G D K Y N S M E D A K V Y  
 225 234 243 252 261 270  
 GTG GCT AAA GTG GAC TGC ACG GCC CAC TCC GAC GTG TGC TCC GCC CAG GGG GTG  
 V A K V D C T A H S D V C S A Q G V  
 279 288 297 306 315 324  
 CGA GGA TAC CCC ACC TTA AAG CTT TTC AAG CCA GGC CAA GAA GCT GTG AAG TAC  
 R G Y P T L K L F K P G Q E A V K Y  
 333 342 351 360 369 378  
 CAG GGT CCT CGG GAC TTC CAG ACA CTG GAA AAC TGG ATG CTG CAG ACA CTG AAC  
 Q G P R D F Q T L E N W M L Q T L N

FIGURE 2A

387	GAG CCA GTG ACA CCA GAG CCG GAA GTG GAA CCG CCC AGT GCC CCC GAG CTC	432
	E P V T P E P E V E P P P S A P E L	
441	AAG CAA GGG CTG TAT GAG CTC TCA GCA AGC AAC TTT GAG CTG CAC GTT GCA CAA	486
	K Q G L Y E L S A S N F E L H V A Q	
495	GAC CAC TTT ATC AAG TTC TTC GCT CCG TGG TGT GGT CAC TGC AAA GCC CTG	540
	G D H F I K F A P W C G H C K A L	
549	CCA ACC TGG GAG CAG CTG GCT CTG GGC CTT GAA CAT TCC GAA ACT GTC AAG	594
	A P T W E Q L A L G L E H S E T V K	
603	AAG GTT GAT TGT ACA CAG CAC TAT GAA CTC TGC TCC GGA AAC CAG GTT	648
	I G K V D C T Q H Y E L C S G N Q V	
657	TAT CCC ACT CTT CTC TGG TTC CGA GAT GGG AAA AAG GTG GAT CAG TAC	702
	R G Y P T L L W F R D G K K V D Q Y	
711	AAG GGA AAG CGG GAT TTG GAG TCA CTG AGG GAG TAC GTG GAG TCG CAG CTG CAG	756
	K G K R D L L E S L R E Y V E S Q L Q	

FIGURE 2B

765 774 783 792 801 810  
CGC ACA GAG ACT GGA GCG ACC GAG AC GTC ACC CCC TCA GAG GCC CCG GTG CTG  
R T E T G A T E T V T P S E A P V L

819 828 837 846 855 864  
GCA GCT GAG CCC GAG GCT GAC AAG GGC ACT GTG TTG GCA CTC ACT GAA AAT AAC  
A A E P E A D K G T V L A L T E N N

873 882 891 900 909 918  
TTC GAT GAC ACC ATT GCA GAA GGA ATA ACC TTC ATC AAG TTT TAT GCT CCA TGG  
F D D T I A E G I T F I K F Y A P W

927 936 945 954 963 972  
TGT GGT CAT TGT AAG ACT CTG GCT CCT ACT TGG GAG GAA CTC TCT AAA AAG GAA  
C G H C K T L A P T W E E L S K K E

981 990 999 1008 1017 1026  
TTC CCT GGT CTG GCG GGG GTC AAG ATC GCC GAA GTA GAC TGC ACT GCT GAA CGG  
F P G L A G V K I A E V D C T A E R

1035 1044 1053 1062 1071 1080  
AAT ATC TGC AGC AAG TAT TCG GTA CGA GGC TAC CCC ACG TTA TTG CTT TTC CGA  
N I C S K Y S V R G Y P T L L L F R

1089 1098 1107 1116 1125 1134  
GGA GGG AAG AAA GTC AGT GAG CAC AGT GGA GGC AGA GAC CTT GAC TCG TTA CAC  
G G K K V S E H S G G R D L D S L H

FIGURE 2C

1143	1152	1161	1170	1179	1188
CGC TTT GTC CTG AGC CAA GCG AAA GAC GAA CTT TAG GAA CAC AGT TGG AGG TCA					
R F V L S Q A K D E L					
1197	1206	1215	1224	1233	1242
CCT CTC CTG CCC AGC TCC CGC ACC CTG CGT TTA GGA GTT CAG TCC CAC AGA GGC					
1251	1260	1269	1278	1287	1296
CAC TGG GTT CCC AGT GGT GGC TGT TCA GAA AGC AGA ACA TAC TAA GCG TGA GGT					
1305	1314	1323	1332	1341	1350
ATC TTC TTT GTG TGT GTG TTT TCC AAG CCA ACA CAC TCT ACA GAT TCT TTA TTA					
1359	1368	1377	1386	1395	1404
AGT TAA GTT TCT CTA AGT AAA TGT GTA ACT CAT GGT CAC TGT GTA AAC ATT TTC					
1413	1422	1431	1440	1449	1458
AGT GGC GAT ATA TCC CCT TTG ACC TTC TCT TGA TGA AAT TTA CAT GGT TTC CTT					
1467	1476	1485	1494	1503	1512
TGA GAC TAA AAT AGC GTT GAG GGA AAT GAA ATT GCT GGA CTA TTT GTG GCT CCT					
1521	1530	1539	1548	1557	1566
GAG TTG AGT GAT TTT GGT GAA AGA AAG CAC ATC CAA AGC ATA GTT TAC CTG CCC					
1575	1584	1593	1602	1611	1620
ACG AGT TCT GGA AAG GTG GCC TTG TGG CAG TAT TGA CGT TCC TCT GAT CTT AAG					

FIGURE 2D

1629 1638 1647 1656 1665 1674  
GTC ACA GTT GAC TCA ATA CTG TGT TGG TCC GTA GCA TGG AGC AGA TTG AAA TGC

1683 1692 1701 1710 1719 1728  
AAA AAC CCA CAC CTC TGG AAG ATA CCT TCA CGG CCG CTG CTG GAG CTT CTG TTG

1737 1746 1755 1764 1773 1782  
CTG TGA ATA CTT CTC TCA GTG TGA GAG GTT AGC CGT GAT GAA AGC AGC GTT ACT

1791 1800 1809 1818 1827 1836  
TCT GAC CGT GCC TGA GTA AGA GAA TGC TGA TGC CAT AAC TTT ATG TGT CGA TAC

1845 1854 1863 1872 1881 1890  
TTG TCA AAT CAG TTA CTG TTC AGG GGA TCC TTC TGT TTC TCA CGG GGT GAA ACA

1899 1908 1917 1926 1935 1944  
TGT CTT TAG TTC CTC ATG TTA ACA CGA AGC CAG AGC CCA CAT GAA CTG TTG GAT

1953 1962 1971 1980 1989 1998  
GTC TTC CTT AGA AAG GGT AGG CAT GGA AAA TTC CAC GAG GCT CAT TCT CAG TAT

2007 2016 2025 2034 2043 2052  
CTC ATT AAC TCA TTG AAA GAT TCC AGT TGT ATT TGT CAC CTG GGG TGA CAA GAC

2061 2070 2079 2088 2097 2106  
CAG ACA GGC TTT CCC AGG CCT GGG TAT CCA GGG AGG CTC TGC AGC CCT GCT GAA

FIGURE 2E

09/744197



2115 2124 2133 2142 2151 2160  
GGG CCC TAA CTA GAG TTC TAG AGT TTC TGA TTC TGT TTC TCA GTA GTC CTT TTA

2169 2178 2187 2196 2205 2214  
GAG GCT TGC TAT ACT TGG TCT GCT TCA AGG AGG TCG ACC TTC TAA TGT ATG AAG

2223 2232 2241 2250 2259 2268  
AAT GGG ATG CAT TTG ATC TCA AGA CCA AAG ACA GAT GTC AGT GGG CTG CTC TGG

2277 2286 2295 2304 2313 2322  
CCC TGG TGT GCA CGG CTG TGG CAG CTG TTG ATG CCA GTG TCC TCT AAC TCA TGC

2331 2340 2349 2358 2367 2376  
TGT CCT TGT GAT TAA ACA CCT CTA TCT CCC TTG GGA ATA AGC ACA TAC AGG CTT

2385 2394 2403 2412 2421 2430  
AAG CTC TAA GAT AGA TAG GTG TTT GTC CTT TTA CCA TCG AGC TAC TTC CCA TAA

2439 2448 2457 2466 2475 2484  
TAA CCA CTT TGC ATC CAA CAC TCT TCA CCC ACC TCC CAT ACG CAA GGG GAT GTG

2493 2502 2511 2520 2529 2538  
GAT ACT TGG CCC AAA GTA ACT GGT GGT AGG AAT CTT AGA AAC AAG ACC ACT TAT

2547 2556 2565 2574 2583 2592  
ACT GTC TGT CTG AGG CAG AAG ATA ACA GCA GCA TCT CGA CCA GCC TCT GCC TTA

FIGURE 2F

09/144197

2601	2610	2619	2628	2637	2646
AAG GAA ATC TTT ATT AAT CAC GTA TGG TTC ACA GAT AAT TCT TTT TTT AAA AAA					
2655	2664	2673	2682	2691	2700
ACC CAA CCC CCT AGA GAA GCA CAA CTG TCA AGA GTC TTG TAC ACA CAA CTT CAG					
2709	2718	2727	2736	2745	2754
CTT TGC ATC ACG AGT CTT GTA TTC CAA GAA AAT CAA AGT GGT ACA ATT TGT TTG					
2763	2772	2781	2790	2799	
TTT ACA CTA TGA TAC TTT CTA AAT AAA CTC TTT TTT TTT AAA AAA AAA A 3'					

FIGURE 2G

**Docket No.: PF-0564 USN**

## DECLARATION AND POWER OF ATTORNEY FOR UNITED STATES PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name, and

I believe that I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if more than one name is listed below) of the subject matter which is claimed and for which a United States patent is sought on the invention entitled

## HUMAN CALCIUM REGULATORY PROTEIN

the specification of which:

       / is attached hereto.

/X/ was filed on January 16, 2001 as application Serial No. 09/744,197 and if this box contains an X / /, was amended on \_\_\_\_\_.

/X/ was filed as Patent Cooperation Treaty international application No. PCT/US99/12385 on July 19, 1999, if this box contains an X / /, was amended on under Patent Cooperation Treaty Article 19 on            2001, and if this box contains an X / /, was amended on           .

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge my duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

I hereby claim the benefit under Title 35, United States Code, §119 or §365(a)-(b) of any foreign application(s) for patent or inventor's certificate indicated below and of any Patent Cooperation Treaty international applications(s) designating at least one country other than the United States indicated below and have also identified below any foreign application(s) for patent or inventor's certificate and Patent Cooperation Treaty international application(s) designating at least one country other than the United States for the same subject matter and having a filing date before that of the application for said subject matter the priority of which is claimed:

**Docket No.: PF-0564 USN**

Country	Number	Filing Date	Priority Claimed
_____	_____	_____	// Yes // No
_____	_____	_____	// Yes // No

I hereby claim the benefit under Title 35, United States Code, §119(e) of any United States provisional application(s) listed below.

Application	Status (Pending,
Serial No.	Abandoned, Patented)
<u>60/160/076</u>	<u>Filed July 20, 1998 Expired</u>

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in said prior application(s) in the manner required by the first paragraph of Title 35, United States Code §112, I acknowledge my duty to disclose material information as defined in Title 37 Code of Federal Regulations, §1.56(a) which occurred between the filing date(s) of the prior application(s) and the national or Patent Cooperation Treaty international filing date of this application:

Application	Status (Pending,
Serial No.	Abandoned, Patented)
_____	_____

I hereby appoint the following:

Lucy J. Billings	Reg. No. <u>36,749</u>
Michael C. Cerrone	Reg. No. <u>39,132</u>
Diana Hamlet-Cox	Reg. No. <u>33,302</u>
Richard C. Ekstrom	Reg. No. <u>37,027</u>
Barrie D. Greene	Reg. No. <u>46,740</u>
Lynn E. Murry	Reg. No. <u>42,918</u>
Shirley A. Recipon	Reg. No. <u>47,016</u>
Susan K. Sather	Reg. No. <u>44,316</u>
Michelle M. Stempien	Reg. No. <u>41,327</u>
David G. Streeter	Reg. No. <u>43,168</u>

10

respectively and individually, as my patent attorneys and/or agents, with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith. Please address all communications to:

Docket No.: PF-0564 USN

**LEGAL DEPARTMENT**  
**INCYTE GENOMICS, INC.**  
**3160 PORTER DRIVE, PALO ALTO, CA 94304**

TEL: 650-855-0555 FAX: 650-849-8886 or 650-845-4166

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Sole Inventor or  
First Joint Inventor: 1-w

Full name:

OLGA BANDMAN

Signature:

Olga Bandman

Date:

16 February, 2001

Citizenship

United States of America

Residence:

Mountain View, California

P.O. Address:

366 Anna Avenue

Mountain View, California 94043

Second Joint Inventor: 2-w

Full name:

Y. TOM TANG

Signature:

Y. Tom Tang

Date:

February 27, 2001

Citizenship

People's Republic of China USA

Residence:

San Jose, California

P.O. Address:

4230 Ranwick Court

San Jose, California 95118

Third Joint Inventor:

3a

Full name:

NEIL C. CORLEY

Signature:

*Neil C. Corley*

Date:

JANUARY 31, 2001

Citizenship

United States of America

Residence:

Castro Valley, California CA

P.O. Address:

20426 Crow Creek Road

Castro Valley, California 94552

Fourth Joint Inventor:

4a

Full name:

KARL J. GUEGLER

Signature:

*Karl J. Gugler*

Date:

02/02, 2001

Citizenship

Switzerland

Residence:

Menlo Park, California CA

P.O. Address:

1048 Oakland Avenue

Menlo Park, California 94025

Fifth Joint Inventor:

5a

Full name:

MARIAH R. BAUGHN

Signature:

*Mariah R. Baughn*

Date:

February 12, 2001

Citizenship

United States of America

Residence:

San Leandro, California CA

P.O. Address:

14244 Santiago Road

San Leandro, California 94577

Docket No.: PF-0564 USN

Sixth Joint Inventor:

*(e-w)*

Full name:

JUNMING YANG

Signature:

*J. Yang*

Date:

February 16, 2001

Citizenship

China

Residence:

San Jose, California

CA

P.O. Address:

7125 Bark Lane

San Jose, California 95129

## SEQUENCE LISTING

<110> INCYTE PHARMACEUTICALS, INC.  
BANDMAN, Olga  
TANG, Y. Tom  
CORLEY, Neil C.  
GUEGLER, Karl J.  
BAUGHN, Mariah R.  
YANG, Junming

&lt;120&gt; CALCIUM REGULATORY PROTEINS

<130> PF-0564 PCT

<140> To Be Assigned

<141> Herewith

<150> 09/119,547

<151> 1998-07-20

<160> 4

<170> PERL program

 $\langle 210 \rangle$  1

<211> 147

&lt;212&gt; PRT

<213> Homo sapiens

 $\langle 220 \rangle$ 

<221> misc feature

<223> Incyte Clone No: 1356311

 $\langle 300 \rangle$ 

<400> 1

Met	Gly	Gly	Cys	Met	His	Ser	Thr	Gln	Asp	Lys	Ser	Leu	His	Leu
1				5					10					15
Glu	Gly	Asp	Pro	Asn	Pro	Ser	Ala	Ala	Pro	Thr	Ser	Thr	Cys	Ala
				20					25					30
Pro	Arg	Lys	Met	Pro	Lys	Arg	Ile	Ser	Ile	Ser	Lys	Gln	Leu	Ala
				35					40					45
Ser	Val	Lys	Ala	Leu	Arg	Lys	Cys	Ser	Asp	Leu	Glu	Lys	Ala	Ile
				50					55					60
Ala	Thr	Thr	Ala	Leu	Ile	Phe	Arg	Asn	Ser	Ser	Asp	Ser	Asp	Gly
				65					70					75
Lys	Leu	Glu	Lys	Ala	Ile	Ala	Lys	Asp	Leu	Leu	Gln	Thr	Gln	Phe
				80					85					90
Arg	Asn	Phe	Ala	Glu	Gly	Gln	Glu	Thr	Lys	Pro	Lys	Tyr	Arg	Glu
				95					100					105
Ile	Leu	Ser	Glu	Leu	Asp	Glu	His	Thr	Glu	Asn	Lys	Leu	Asp	Phe
				110					115					120
Glu	Asp	Phe	Met	Ile	Leu	Leu	Leu	Ser	Ile	Thr	Val	Met	Ser	Asp
				125					130					135
Leu	Leu	Gln	Asn	Ile	Arg	Asn	Val	Lys	Ile	Met	Lys			
				140					145					

 $\langle 210 \rangle$  2

<211> 363

<212> PRT







WO 00/05368

PCT/US99/12385

ctctcagtg	gagaggttag	ccgtgatgaa	agcagcgta	cttctgaccg	tgcctgagta	1800
agagaatgct	gatgccataa	ctttatgtgt	cgatacttgt	caaatacagtt	actgttcagg	1860
ggatccttct	gtttctcacg	gggtgaaaca	tgtcttttagt	tcctcatgtt	aacacgaagc	1920
cagagcccac	atgaactgtt	ggatgtcttc	cttagaaagg	gtaggcacgt	aaaattccac	1980
gaggctcatt	ctcagtatct	cattaactca	ttgaaagatt	ccagttgtat	ttgtcacctg	2040
gggtgacaag	accagacagg	ctttcccagg	cctgggtatc	cagggaggct	ctgcagecct	2100
gctgaagggc	cctaactaga	gttctagagt	ttctgattct	gtttctcagt	agtcctttta	2160
gaggcttgct	atacttggtc	tgcttcaagg	aggctcgacct	tctaattgtat	gaagaatggg	2220
atgcatttga	tctcaagacc	aaagacagat	gtcagtgggc	tgctctggcc	ctgggtgtgca	2280
cggctgtggc	agctgttgat	gccagtggtc	tctaactcat	gctgtccttg	tgattaaaca	2340
cctctatctc	ccttgggaat	aagcacatac	aggcttaagc	tctaagatag	ataggtgttt	2400
gtcctttttac	catcgagcta	cttcccataa	taaccacttt	gcataccaaca	ctcttcaccc	2460
acctcccata	cgcaagggga	tgtggatact	tggcccaaag	taactgggtg	taggaatctt	2520
agaaacaaga	ccacttatac	tgtctgtctg	aggcagaaga	taacagcagc	atctcgacca	2580
gcctctgcct	taaaggaaat	ctttattaat	cacgtatggg	tcacagataa	ttcttttttt	2640
aaaaaaaccc	aacccccctag	agaagcacia	ctgtcaagag	tcttgtagac	acaacttcag	2700
ctttgcatca	cgagtcttgt	attccaagaa	aatcaaagtg	gtacaatttg	tttgtttaca	2760
ctatgatact	ttctaaataa	actctttttt	tttaaaaaaa	aaa		2803